On using virtual circuits for GridFTP transfers

Z. Liu*, M. Veeraraghavan*, Z. Yan*, C. Tracy†, J. Tie‡, I. Foster‡, J. Dennis§, J. Hick¶, Y. Li‖ and W. Yang‖
*University of Virginia (UVA), Charlottesville, VA 22904, {zl4ef,mv5g,zy4d}@virginia.edu
†Energy Sciences Network (ESnet), Berkeley, CA 94720, ctracy@es.net
‡University of Chicago, Chicago, IL 60637, {jtie,foster}@cs.uchicago.edu
§National Center for Atmospheric Research (NCAR), Boulder, CO 80305, dennis@ucar.edu
¶National Energy Research Scientific Computing Center (NERSC), Berkeley, CA 94720, jhick@lbl.gov
‖SLAC National Accelerator Laboratory, Menlo Park, CA 94025 {ytl, yangw}@slac.stanford.edu
**Argonne National Laboratory, Argonne, IL 60439, foster@anl.gov

***Abstract*—The goal of this work is to characterize scientific data transfers and to determine the suitability of dynamic virtual circuit service for these transfers instead of the currently used IP-routed service. Specifically, logs collected by servers executing a commonly used scientific data transfer application, GridFTP, are obtained from three US super-computing/scientific research centers, NERSC, SLAC, and NCAR, and analyzed. Dynamic virtual circuit (VC) service, a relatively new offering from providers such as ESnet and Internet2, allows for the selection of a path on which a rate-guaranteed connection is established prior to data transfer. Given VC setup overhead, the first analysis of the GridFTP transfer logs characterizes the duration of sessions, where a session consists of multiple back-to-back transfers executed in batch mode between the same two GridFTP servers. Of the NCAR-NICS sessions analyzed, 56% of all sessions (90% of all transfers) would have been long enough to be served with dynamic VC service. An analysis of transfer logs across four paths, NCAR-NICS, SLAC-BNL, NERSC-ORNL and NERSC-ANL, shows significant throughput variance, where NICS, BNL, ORNL, and ANL are other US national laboratories. For example, on the NERSC-ORNL path, the inter-quartile range was 695 Mbps, with a maximum value of 3.64 Gbps and a minimum value of 758 Mbps. An analysis of the impact of various factors that are potential causes of this variance is also presented.**

I. Introduction

Large scientific data sets are being created in a number of ways: by instruments (e.g., telescopes), through experiments (e.g., Large Hadron Collider (LHC) experiments), through simulations on high-performance computing platforms, and from model reanalyses [1]. These data sets are typically stored at the institutions where they are created.

Geographically distributed scientists then download these data sets to their local computer clusters. Several applications, such as GridFTP [2], FDT [3], bbFTP [4], `scp` and `sftp`, are used for these data transfers. For large data transfers, throughput is an important measure as it determines the total time required for the transfers. To achieve high throughput, scientific users often invest in high-end data transfer servers with high-speed disk arrays and wide-area network (WAN) access links. With such equipment, scientists can move files end-to-end at multiple, e.g, 2-4, Gbps, which is a significant fraction of the typical core network link capacity, which is 10 Gbps. The TCP flows created by such high-speed transfers of large data sets were termed α flows by Sarvotham et al. [5] as they dominate flows from general-purpose Internet applications. These α flows are responsible for increasing the burstiness of IP traffic [5].

For operational reasons, research-and-education network providers, such as ESnet, are interested in carrying these α flows on virtual circuits rather than on their IP-routed paths. On the *positive* side, virtual circuits have the potential for reducing throughput variance for the large data transfers as they can be provisioned with rate guarantees, a feature not available with IP-routed service. Second, since virtual circuits require a setup procedure prior to use, there is an opportunity for a management software system such as On-Demand Secure Circuits and Advance Reservation System (OSCARS) [6] to explicitly select a path for the virtual circuit based on current network conditions, policies, and service level agreements. Third, as part of the setup procedure, packet classifiers on the input side and packet schedulers on the output side of router interfaces can be configured to isolate α-flow packets into their own virtual queues. Such configurations will prevent packets of general-purpose flows from getting stuck behind a large-sized burst of packets from an α flow. The result is a reduction in delay variance (jitter) for the general-purpose flows. On the *negative* side, virtual circuits incur the overhead of setup delay. If the duration of an α flow is not significantly longer than setup delay, only static virtual circuits can be used, which can potentially be used in an intra-domain context, but is an unscalable solution for inter-domain usage.

The *problem statement* of this work is to analyze GridFTP transfer logs, which capture usage patterns of scientific data movement, to determine the feasibility and value of using dynamic virtual circuits. Our *key findings* that advance the overall state of understanding about scientific data movement are as follows: **(i)** Dynamic VCs can be used, in spite of their setup delay overhead, for a considerable portion of the GridFTP transfers (e.g., 78.4% of all SLAC-BNL transfers). Before the analysis, our hypothesis was that most transfers are short-lived, and with increasing link rates, a very small percentage of transfers will last long enough to justify the VC setup delay overhead, which is currently 1 min in the ESnet implementation. But data analysis showed that most transfers are part of sessions in which users execute multiple back-to-back transfers to the same remote server, which then

SC12, November 10-16, 2012, Salt Lake City, Utah, USA
978-1-4673-0806-9/12/$31.00 ©2012 IEEE

makes the total size of the data transferred large enough that even under high-rate assumptions, the transfer duration will be considerably longer than VC setup delay. **(ii)** The highest observed GridFTP transfer throughput is $\sim$4.3 Gbps, but on all four paths for which data was analyzed, transfers have been observed at values as high as 2.5 Gbps. Thus, these transfers do consume a significant portion of link capacity, which is typically 10 Gbps on these paths, and hence these α flows have the potential for adding delays to packets of other flows. While it was previously known that scientific transfers can reach high rates, the value of this analysis lies in providing the actual observed rates. **(iii)** The aggregate throughput of 8 TCP-stream transfers is higher than that of 1 TCP-stream transfers for small files, but not for large files. The implication of not seeing a significant difference in throughput between 8-stream transfers and 1-stream transfers for large files is that packet losses are rare. This finding can be taken into account while designing new transport protocols for high bandwidth-delay product paths for the scientific community. **(iv)** The backbone network links are relatively lightly loaded while science flows dominate the total traffic. This finding is based on a correlation analysis between GridFTP transfer logs and Simple Network Management Protocol (SNMP) link usage data. The results of this analysis were surprising in the sense that the correlation coefficients were much higher than expected especially as the links were ESnet backbone links. Our expectation was that on backbone links, aggregated traffic from general-purpose flows will dominate, but this analysis shows that individual GridFTP transfers dominate the total number of bytes on backbone links. **(v)** Variance in transfer throughput has been a major concern to users. But our findings are pointing to competition for server resources rather than network resources. For example, we found that the throughput of a particular transfer is affected by concurrent transfers being served by the same data transfer node. Since link utilizations are low, this effect is more likely due to competition for CPU and disk I/O resources. This means solutions to reduce throughput variance require scheduling of server resources prior to data transfers, not just network bandwidth.

Section II provides background information on GridFTP and virtual circuit services. Section III reviews related work. Section IV describes aspects of VC service relevant to large dataset movement. The analyzed GridFTP datasets are described in Section V. Sections VI and VII present the results of our analysis of GridFTP transfer logs on four paths, and discuss the implications of our findings on the usage of dynamic VCs. The paper is concluded in Section VIII.

II. Background

GridFTP GridFTP [2] is a set of extensions to the File Transfer Protocol (FTP) [7] for high-rate data transfers. Mechanisms used to increase throughput include streaming (the use of parallel TCP flows) and striping (data blocks stored on multiple computers at one end are transferred in parallel to multiple computers at the other end). Third-party transfers, recovery from failures during transfers, and security support (authentication, integrity and privacy) are other features that enable GridFTP to offer users secure and reliable data transfers. The Globus organization [8], [9] and others have implemented open-source Globus GridFTP clients and servers. This application is widely used in the scientific community.

The Globus GridFTP server supports usage statistics collection. GridFTP servers send usage statistics in UDP packets at the end of each transfer to a server maintained by the Globus organization. Administrators of GridFTP servers have the option to disable this feature. For each transfer, the following information is logged: transfer type (store or retrieve), size in bytes, start time of the transfer, transfer duration, IP address and domain name of the GridFTP server, number of parallel TCP streams, number of stripes, TCP buffer size, and block size. Importantly, the IP address/domain name of the other end of the transfer is not listed for privacy reasons. GridFTP servers also maintain local logs of these transfers.

Virtual circuit service Both research-and-education network (REN) providers, such as ESnet, Internet2, GEANT2, JGN2plus, and commercial providers such as AT&T and Verizon, offer optical circuit services in addition to their basic IP-routed service [10]. Two types of circuit services are offered: *static* (also called leased lines), and *dynamic*. With a static circuit service, a customer specifies the endpoints, rate, and duration, and a single contract is issued to the customer by the provider for a static circuit. Typically, static circuit service durations are on the order of months. With a dynamic circuit service, a customer purchases a high-speed access link to the provider's network on a long-term contract (e.g., two years). Subsequently, short-duration circuits of rate less than or equal to the customer's access link rate can be requested to an endpoint in any other customer network, which also has an access link for dynamic circuit services. As an example of a dynamic circuit service, consider a wireless cellular/wired phone service. This service requires customers to first sign a long-term contract with a provider, which then allows them to place calls to other customers.

There are two differences between phone service and high-speed optical dynamic circuit service. First, the phone service allows for users to request circuits to customers of other providers, i.e., inter-domain service is supported. Commercial high-speed optical dynamic circuit services are currently only intra-domain, but REN providers are experimenting with inter-domain service. This inter-domain capability is required as more campus and regional networks are starting to offer high-speed dynamic circuit services through projects such as Dynamic Network System (DYNES) [11]. Second, phone service only supports calls that request connectivity for immediate usage, while high-speed optical dynamic circuit service supports advance reservations. Specifically, ESnet and Internet2 deploy Inter-Domain Controller Protocol (IDCP) [12] schedulers that receive and process advance-reservation requests for virtual circuits of specified rate and duration, provision these circuits at the start of their scheduled times, and release the circuits

when their durations end. Such advance-reservation service is required when the requested circuit rate is a significant portion of link capacity if the network is to be operated at high utilization and with low call blocking probability [13].

III. Related Work

Experimental work has been done to compare different applications for high-speed file transfers [14], [15]. Measurements obtained from the experiments are analyzed. In contrast, our work analyzes data collected for actual transfers executed by scientific users.

There are several papers [16], [17] on Grid monitoring systems. But to our knowledge, there is no other published work that analyzes usage statistics collected by deployed GridFTP servers, which is the focus of this work.

Finally, relevant to this work are flow classification papers. Lan and Heidemann [18] classify flows on four dimensions: size (bytes), duration, throughput, and burstiness, and report that 68% of `porcupine` (high burstiness) flows in an analyzed data set were also `elephant` (large sized) flows. Sarvotham et al. [5] report that a "majority of burst causing connections are due to large file transfers over a large bottleneck bandwidth end-to-end path," which as mentioned earlier are referred to as α flows.

IV. Use of virtual circuits for data transfers

As stated in Section I, there are both positives and negatives in the use of virtual circuits for large dataset movement. There are two dimensions of virtual circuits to consider: (i) intra- vs. inter-domain, and (ii) static vs. dynamic.

Intra-domain virtual circuit (VC) service is easier to deploy as it involves just a single provider. The question is whether such a service can be of value for large dataset movement. Of the positives listed in Section I, the second and third positives whereby the provider can control the path taken by the flows, and can isolate these α flows from general-purpose flows, can be realized with intra-domain VCs. With automatic α flow identification [19], packets from α flows can be redirected to intra-domain VCs, such as MPLS label switched paths, that have been preconfigured between ingress-egress router pairs. These VCs can be static if the provider network does not have a large number of routers. Alternatively, solutions such as Lambdastation [20] can be used to have end user applications, which generate large-sized high-speed transfers, signal their intention (before starting their transfers) to network management systems deployed within a provider's network. Such notifications would allow the network management systems to configure the redirection of α flows to static intra-domain VCs (using firewall filters), and even allow for dynamic intra-domain VC setup in large provider networks.

To realize the first positive of VCs noted in Section I, which is a reduction of throughput variance, VCs have to be end-to-end, and hence in many cases inter-domain. Also, it is often important to control the inter-domain path. With today's IP-routed service, a tier-2 provider has little control over the path taken by an incoming α flow, which is influenced by the Border Gateway Protocol (BGP) configurations in the campus and regional networks of the data transfer source. Say, for example, a tier-2 provider has purchased transit service with a certain service level agreement (SLA), e.g., 300 Mbps, from a tier-1 provider, and the path taken by the incoming α flow chooses this transit link. Since α flows are high-rate flows, the rate of a single α flow could exceed the SLA transit rate. This would result in the tier-2 provider having to pay additional fees to its tier-1 provider. Therefore, providers want to control the path taken by inter-domain α flows. In an inter-domain context, the static circuit solution does not scale as the number of providers increases. Therefore, dynamic circuit service is required.

In current implementations of circuit schedulers such as the OSCARS Inter-Domain Controller (IDC) for ESnet, VC setup delay is minimally 1 min because the IDC is designed primarily for advance reservations. Users or applications send a `createReservation` message to the IDC with the following parameters: `startTime`, `endTime`, `bandwidth`, and `circuit endpoint addresses`. Two options are provided for the circuit provisioning phase, *automatic signaling*, in which the IDC automatically sends a request to the ingress router to initiate circuit provisioning just before the `startTime` of the circuit, or *message signaling*, in which the user or application generates an explicit `createPath` message to trigger circuit provisioning. Commonly, the former technique is used. The IDC has the opportunity to collect all provisioning requests that start in the next minute and send them in batch mode to the ingress router. This solution however results in a minimum 1-min VC setup delay if a data transfer application sends a VC setup request to the IDC for immediate usage. Since scientific users are accustomed to initiating GridFTP or other transfer applications on an as-needed basis rather than on an advance-scheduled basis, this VC setup delay overhead is important when considering the use of dynamic VC service for data movement.

V. Data Sets

GridFTP usage logs were obtained from NERSC, NCAR and SLAC. In general, these data sets can be obtained from the Globus organization with the approval of the enterprise running the GridFTP servers, or directly from the enterprises themselves. We used both methods for this data procurement. From these sets of logs, transfers on four pairs of paths were analyzed: (i) NERSC-ORNL (Sep. 2010), (ii) NERSC-ANL (Mar. 4 - Apr. 22, 2012), (iii) NCAR-NICS (2009-2011), and (iv) SLAC-BNL (Feb. 10 - Apr. 26, 2012). In some cases, the periods chosen were based on availability and in others the choice was random. Future data sets may be more easily obtained from Globus Online [21].

Two terms are used in this analysis: transfers and sessions. The term `transfer` refers to a single entry in the GridFTP transfer logs, which corresponds to a single file. The term `session` refers to multiple transfers executed in batch mode by an automated script. A configurable parameter, g, is used to set the maximum allowed gap between the end of one transfer

and the start of the next transfer within a session. The gap between end of one transfer and the start of the next transfer in a log could be negative as multiple transfers can be started concurrently. Such transfers are part of the same session.

The NCAR and SLAC GridFTP logs include information about the remote IP address, which allowed for an analysis of session characteristics. Unfortunately, in the NERSC data set, the remote IP address was anonymized for privacy reasons. Without knowledge of the remote end of the transfers, the NERSC GridFTP transfers could not be grouped into sessions. However, it was possible to isolate GridFTP transfers corresponding to periodic administration-run tests from NERSC to ORNL and NERSC to ANL, for which transfer throughput analysis was carried out. The results of these analyses are presented in Section VI. In Section VII, the importance of different factors on throughput variance is analyzed by combining the GridFTP data sets with data such as SNMP link usage measurements.

VI. Characterizing sessions and transfers

An analysis of the size and duration of sessions is presented in Section VI-A, which also includes an analysis of transfer throughput for two of the four paths listed in Section V. Additional transfer throughput analyses are presented for the two other paths in Section VI-B.

A. Session analysis

Often scientists move lots of files because their simulation programs or experiments create many files. Scripts are used to have GridFTP move all files in one or more directories. The GridFTP usage logger logs information for each file, as described in Section II. Each such file movement is referred to as a "transfer," and "sessions" consist of one-or-more transfers, as defined in Section V. Since a virtual circuit (VC), once established, can be used for all transfers within a session before VC release, the question considered in this analysis is whether or not durations of sessions are long enough to justify the VC setup delay overhead.

In grouping transfers into sessions by examining the GridFTP logs, a value needs to be selected for the configurable parameter g, defined in Section V. Since VC setup delay in today's ESnet deployment is approximately 1 min, values of 1 min and 2 min are considered for this g parameter in the analysis below. If VC release takes 1 min as well, then it seems reasonable to hold the VC even if there is a 2-min gap between the end of one transfer and the start of the next transfer. Unlike with circuits, a virtual circuit allows for shared usage of assigned capacity, in that packets from other VCs can be served during idle periods. Therefore, to some extent, holding a VC open even when idle is not an expensive proposition. On the other hand, VCs add to administrative overhead, and hence should not be held open indefinitely. Therefore we use these two values of 1 min and 2 min for g in the analysis presented below. A policy of no idle periods on a VC would require a g value of 0, which is also included in the analysis.

TABLE I: NCAR-NICS sessions and transfers; $g = 1$ min

Characterization of session sizes, in MB

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8,793 (bytes)	5,808.7	70,708.4	263,771.4	320,600	2,873,868.5

Characterization of session durations, in s

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.05	210.5	1,445	4,039	5,261	48,420

Characterization of transfer throughput, in Mbps

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.1 (bps)	298	468.3	506.1	682.2	4,227

TABLE II: SLAC-BNL sessions and transfers; $g = 1$ min

Characterization of session sizes, in MB

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
812 (bytes)	273	1,195	24,045	4,860	12,037,604

Characterization of session durations, in s

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.05	18.95	58.92	315.9	151.1	95,080

Characterization of transfer throughput, in Mbps

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.003	22.57	112.8	130.4	183.1	2,560

Two data sets are used for this session analysis: (i) NCAR-NICS data set, which consists of $52,454$ transfers, and (ii) the SLAC-BNL data set, which consists of $1,021,999$ transfers. Tables I and II show the size and duration statistics about *sessions*, but throughput statistics about *transfers*, for the two data sets, respectively, assuming g is 1 min. The total size of a session and the duration of a session are relevant in determining the feasibility of using dynamic virtual circuits. However, for the throughput measure, transfer throughput is analyzed rather than session throughput. This is because session throughputs could be lower if some of the individual transfers within a session had lower throughput.

Some points to observe in the results reported in Tables I and II are as follows. The maximum transfer throughput in the SLAC-BNL data set is lower at 2.56 Gbps when compared to the maximum throughput on the shorter NCAR-NICS path at 4.23 Gbps. The distributions of the session size for both data sets are skewed right, e.g., in the SLAC-BNL data set, the median ($\sim$1.1 GB) is significantly smaller than its mean ($\sim$24 GB). The largest session of size $\sim$12 TB in the SLAC-BNL dataset took 26 hours and 24 minutes to complete, receiving an effective throughput of 1.06 Gbps. The longest-duration session occurred in the NCAR-NICS data set, with a duration of 13 hours and 27 minutes (48420 sec). The size of this session was 2.4 TB, giving it an effective rate of 410 Mbps. This session throughput is lower than even the third-quartile throughput for the NCAR-NICS path of 682.2 Mbps.

Table III shows the number of different types of sessions under different assumptions for the g parameter. The 1 min value for g appears to offer significant advantages relative to a 0 value, by decreasing the number of single-transfer sessions. This effectively makes it more feasible to use VCs as session durations are likely to be longer than VC setup overhead.

Table IV shows the percentage of sessions that can tolerate

TABLE III: Impact of the g parameter on number of sessions

Data set	g	Number of single-transfer sessions	Number of multi-transfer sessions	Percent with 1 or 2 transfers	Highest number of transfers in a session	Number of sessions with ≥ 100 transfers
NCAR-NICS	$g = 0$	$25,050$	$16,163$	93.2%	702	1
NCAR-NICS	$g = 1$ min	25	186	18.5%	$19,510$	86
NCAR-NICS	$g = 2$ min	14	171	12.9%	$19,510$	86
SLAC-BNL	$g = 0$	$41,765$	$74,133$	53.6%	$9,120$	$1,277$
SLAC-BNL	$g = 1$ min	779	$9,420$	19.3%	$30,153$	$1,412$
SLAC-BNL	$g = 2$ min	358	$5,795$	15.7%	$38,497$	$1,068$

TABLE IV: Percentage of sessions suitable for using VCs (percentage of transfers)

	NCAR data set		SLAC data set	
setup delay	1 min	50 ms	1 min	50 ms
g = 0	0.12% (2.14%)	87.09% (89.33%)	1.95% (39.41%)	52.58% (89.68%)
g = 1 min	56.87% (90.54%)	92.89% (98.04%)	12.54% (78.38%)	93.56% (99.73%)
g = 2 min	62.16% (90.71%)	94.59% (98.17%)	15.93% (85.49%)	94.47% (99.85%)

the dynamic VC setup overhead under different assumptions of the g value, and under two assumptions for VC setup delay. The VC setup delay of 1 min is based on the ESnet implementation, while 50 ms is chosen as the lowest value (round-trip propagation delay across the US) if VC setup message processing is implemented in hardware [22].

Our methodology for determining these percentages is as follows. Instead of considering the actual durations of sessions, which could be high because of other factors such as disk I/O access rates, new hypothetical durations are computed by dividing session sizes by the third quartile of transfer throughput. The question posed is for what percentage of the sessions would the VC setup delay overhead represent one-tenth or less of session durations if the session throughput is assumed to be as high as the third-quartile throughput across all transfers (which would make most session durations lower than their actual durations)? For example, what percentage of the sessions would be longer than 10 min (for the VC setup delay assumption of 1 min) assuming a session throughput value equal to the third-quartile transfer throughput value (682.2 Mbps for the NCAR-NICS data, and 183.1 Mbps for the SLAC-BNL data).

Below is a discussion of the numbers in Table IV corresponding to the $g = 1$ min row. Out of the $211 (= 25 + 186)$ sessions in NCAR-NICS data set (see Table III), 120 sessions would have lasted longer than 10 min if throughput was 682.2 Mbps. This constitutes 56.87% of all sessions. These 120 sessions include 90% of all transfers. For the SLAC-BNL data set, only 12.5% $(1,279)$ of all $(10,199)$ sessions would have lasted longer than 10 min if the throughput was 183.1 Mbps. It is interesting though that these 1,279 sessions include $800,996$ transfers, which represents 78.4% of all transfers.

If VC setup delay can be decreased to say 50 ms, then dynamic VCs can be used for sessions of sizes 42 MB or larger in the NCAR-NICS data set (assuming the same factor of 10 and the same throughput of 682.2 Mbps). Out of the 211 sessions, 196 sessions are $\geq$42MB, which is a high 93% of the 211 sessions observed in the NCAR-NICS data. Under the same assumption, for the SLAC-BNL data set, dynamic VCs can be used for more than 93.6% of all $10,199$ sessions.

The percentages discussed above are smaller under the $g = 0$ assumption, and larger under the more tolerant $g = 2$ min assumption, as seen in Table IV. In *summary*, dynamic VCs can be used, in spite of their setup delay overhead, for a considerable portion of the GridFTP transfers.

TABLE V: The 32GB NERSC-ORNL transfers (145)

	Duration (s)	Throughput (Mbps)
Min	75.4	757.9
1st Qu.	141.2	1251
Median	183.4	1499
Mean	186.6	1625
3rd Qu.	219.7	1947
Max	362.7	3644

B. Transfer throughput

In the previous sub-section, in addition to an analysis of sessions to determine the feasibility of using dynamic virtual circuits, statistics pertaining to transfer throughput were presented in Tables I and II for the NCAR-NICS and SLAC-BNL paths, respectively. In this subsection, transfer throughput statistics are provided for two additional paths: NERSC-ORNL and NERSC-ANL, as noted in Section V. Transfer throughput is important to characterize because of the potential negative impact of α flows on general-purpose flows, especially real-time sensitive audio/video flows.

NERSC-ORNL transfers

From the NERSC logs, a subset of 145 32GB transfers were identified as test transfers between the NERSC GridFTP server and an ORNL GridFTP server. A summary of throughput and duration of these 32GB transfers is presented in Table V. It shows that, even though these transfers occurred across the same path, which means path characteristics such as bottleneck link rate and Round-Trip Time (RTT) are the same for all transfers, there was considerable variance in throughput (the inter-quartile range is 695Mbps).

ANL-NERSC transfers Our dataset comprises a total of 334 test transfers, which—conveniently for our study—are of four different types: memory-to-memory (84), memory-to-disk (78), disk-to-memory (87), and disk-to-disk (85). The statistics on transfer throughput are shown in Table VI. Variance in transfers involving disks was expected to be higher than in memory-to-memory transfers, but this is not the case. In fact,

TABLE VI: Throughput of ANL-NERSC transfers (Mbps)

	mem-mem	mem-disk	disk-mem	disk-disk
Min	308.9	202.4	297.4	10.85
1st Qu.	1149	599.6	1265	527.3
Median	1472	819.0	1569	645.9
Mean	1463	789.6	1563	670
3rd Qu.	1772	1007	1851	841.3
Max	2706	1354	2529	1079
CV	35.69%	31.63%	30.80%	33.10%

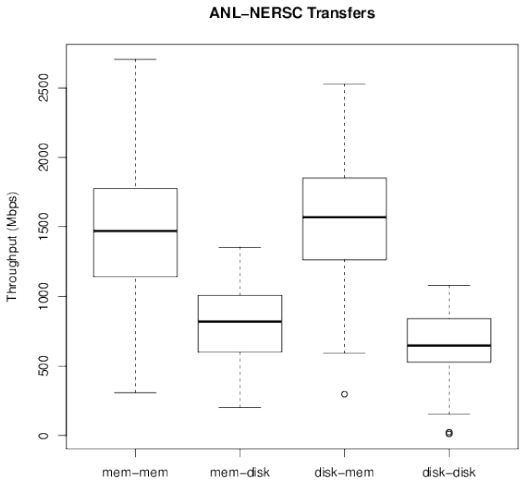

Fig. 1: Throughput variance for ANL-to-NERSC transfers

the coefficient of variation (CV) is highest for memory-to-memory transfers.

Fig. 1 shows box plots for the four categories. It shows that the NERSC disk I/O system is a bottleneck because memory-to-disk transfers and disk-to-disk transfers show lower median throughput than memory-to-memory or disk-to-memory transfers from ANL to NERSC.

In addtition to the observations about high variance in transfer throughput, we also observe that these transfers do consume a significant portion of link capacity, which is typically 10 Gbps on these paths. Among the four paths, the highest throughput was observed on the NCAR-NICS path at $\sim$4.3 Gbps, but on all four paths, transfers have been observed at 2.5 Gbps, hence these α flows have the potential for adding delays to packets of other flows.

VII. Analyzing the impact of different factors on transfer throughput

Several factors could be responsible for the observed variance of throughput between two specific GridFTP servers across the same network path. These include: (i) number of stripes, (ii) number of parallel TCP streams, (iii) time of day, (iv) utilization of network links on the end-to-end path, (v) concurrent GridFTP transfers at the two servers, (vi) packet losses due to router/switch buffer overflows or bit errors, and (vii) CPU usage and I/O usage of the two servers. These are not all independent factors. So far, we have been able to conduct factor analysis for the first five factors. Additional monitoring is required to obtain the data necessary for an analysis of the remaining two factors. The NCAR-NICS dataset is used for factor (i) listed above, the SLAC-BNL dataset is used for the factor (ii), the NERSC-ORNL dataset is used for factors (iii) and (iv), and finally the NERSC-ANL dataset is used to analyze the impact of the factor (v). Unfortunately, only single data sets are available for each of these analyses. The conclusions will be validated further with other data sets in future work.

TABLE VII: Throughput variance of 16GB/4GB transfers in NCAR data set (Unit: Mbps)

	16G	4G
Min.	10.79	4.141
1st Qu.	619	600
Median	838.7	873.4
Mean	834.2	848.5
3rd Qu.	1038	1091
Max.	1543	1587
Std. Dev.	293.43	342.01

This analysis is relevant to the question of using virtual circuits (VCs) for two reasons. First, if the main cause of throughput variance is competing traffic on the network links, then rate-guaranteed VCs will be useful in reducing this variance. The second reason for understanding the causes of throughput variance is to provide a mechanism for the data transfer application to estimate the rate and duration it should specify when requesting a virtual circuit based on values chosen for parameters such as number of stripes, number of streams, etc.

A. Impact of number of stripes

Striped transfers are those in which multiple servers are involved at the two ends [9]. Specifically, the dependence on number of stripes is studied for transfers of two size ranges, [16, 17) GB and [4, 5) GB (henceforth referred to as `16G` and `4G`, respectively), which constitute 87% of the top 5% largest-sized transfers in the NCAR-NICS dataset. The significant throughput variance experienced by these transfers is shown in Table VII. The capacity of the NCAR GridFTP cluster *frost* was reduced from 3 servers in 2009 to 1 server in 2011. In year 2009, the number of servers was either 1 or 3, but in year 2010, it was mostly 2 servers, and in year 2011, it was mostly 1 server. This explains the observed trends in throughput in Table VIII.

An analysis based on *number of stripes* shows the more direct dependence of throughput on number of stripes. A comparison of the throughput in Table IX with those in Table VIII shows the effect of the change in the number of servers in the three years. The minimum and maximum values are not relevant since any single transfer could achieve a low througput or a high throughput based on other factors besides the number of stripes. But the median column is the one to consider. This is higher when the number of stripes is higher as seen for both transfer sets (16 GB and 4 GB) in Table IX.

TABLE VIII: Throughput of 16GB/4GB transfers in NCAR data set (Mbps)

Year based analysis of 16GB transfers								
Year	No. of Transfers	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Standard Deviation
2009	1076	10.79	707.3	889.3	877	1075	1543	294.17
2010	233	95.36	516.5	619.2	651.7	742.9	1150	205.53
2011	12	441.73	480.71	538.77	539.1	575.39	652.07	66.92
Year based analysis of 4GB transfers								
Year	No. of Transfers	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Standard Deviation
2009	853	4.14	593.1	873.1	849.2	1125	1587	366.01
2010	247	72.99	767	977.1	903.1	1083	1209	225.09
2011	37	296.27	376.13	497.81	475.63	556.6	637.13	101.12

TABLE IX: Throughput of 16GB/4GB transfers in NCAR data set (Mbps)

Stripes based analysis of 16GB transfers								
No. of Stripes	No. of Transfers	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Standard Deviation
1	13	441.73	483.7	541.75	546.84	616.48	652.07	69.88
2	547	10.79	542	714	705.4	855.2	1207	212.34
3	761	19.83	748.7	976	931.6	1150	1543	306.96
Stripes based analysis of 4GB transfers								
No. of Stripes	No. of Transfers	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Standard Deviation
1	18	372.2	449.6	506.2	569.2	574.2	1309	225.85
2	447	72.99	566.7	773.1	772.8	1021	1209	245.37
3	759	4.14	625.6	927.6	875.8	1169	1587	375.38

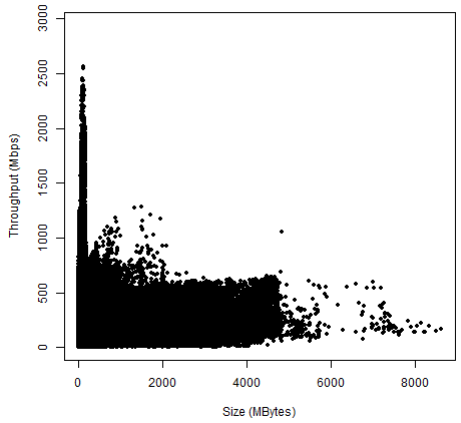

Fig. 2: Throughput of SLAC-BNL transfers

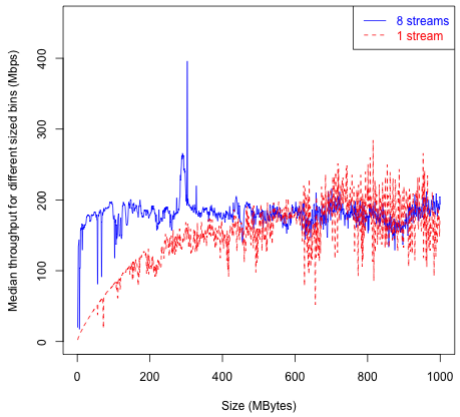

Fig. 3: Throughput of 8-stream and 1-stream transfers of size (0, 1GB)

B. *Impact of number of parallel TCP streams*

The SLAC-BNL dataset, which has 1021999 transfers, is used for this analysis. Fig. 2 plots throughput as a function of file size. There is considerable throughput variance among transfers of the same size. All transfers used a single stripe, and hence the number of stripes is not a factor for this dataset. Of the 1021999 transfers, 864762 (84.615%) transfers consisted of multiple (more than one) parallel TCP streams. A peak value of 2.56 Gbps occurred for a transfer of size 124.6 MB. There were 2215 transfers with throughput greater than 1.5 Gbps, of which 85.37% occurred between 2-3 AM SLAC time on one particular day, Apr. 2, 2012. These transfers were of size 56.73-170.88 MB.

To analyze the effect of the number of parallel TCP streams on GridFTP transfer throughput, transfers were divided, based on their size, into bins. For transfers of size [0 GB, 1 GB], the bin size is chosen to be 1 MB, while for transfers of size (1 GB, 4 GB], the bin size is chosen to be 100 MB. The reason for these bin size selections is to keep the sample size of transfers in each bin to a large enough value for statistical analysis. At larger transfer sizes, there are fewer transfers and hence the bin size is larger. The next step is to partition the transfers in each file size bin into two groups: (i) 1-stream transfers and (ii) 8-stream transfers. The median throughput is computed for each group (1-stream transfers and 8-streams transfers) for each file size bin. The reason for considering median and not mean is to avoid the effects of outliers.

Fig. 3 plots the median transfer throughput for each file size bin in the [0 GB, 1 GB] range. It shows that for small file sizes, the median throughput for the 8-streams group is higher than the median throughput for the 1-streams group. This can be explained with TCP's Slow Start behavior. If for each transfer, the TCP congestion window (`cwnd`) starts at 1

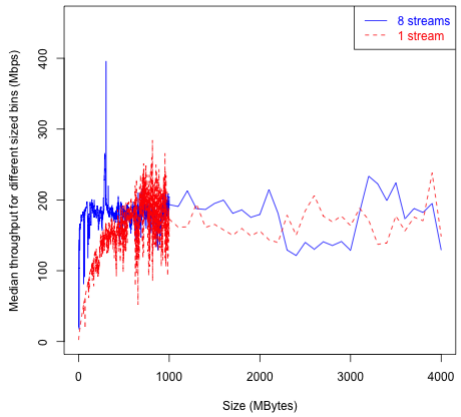

Fig. 4: Throughput of 8-stream and 1-stream transfers of size (0, 4GB)

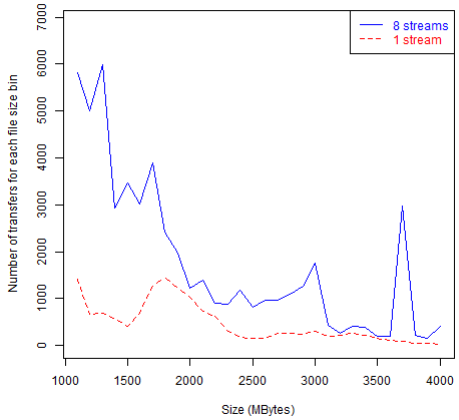

Fig. 5: Number of observations for each file size bin

maximum segment size (MSS), then with 1 TCP connection, the throughput will be lower than with 8 TCP connections. Currently, we do not have a good explanation for the observed spike in median throughput to 400 Mbps for the 8-streams plot for the [302 MB, 303 MB] file size bin in Fig. 3. The number of transfers for this file size bin for the 8-streams plot is 588, which is a reasonably large sample size. The bandwidth-delay product (BDP) for this path is 10 Gbps $\times$ 80 ms = 95.4 MB (assuming 1 MB = 2^{20} bytes). The median throughput is the same, at approximately 200 Mbps for files larger than 575 MB for the 1-stream group, and for files larger than 146 MB for the 8-streams group.

Fig. 4 shows that the median throughput for the 1-stream and 8-stream groups for files larger than 1 GB is roughly the same. The first observation is that the 8-stream plot shows a roughly 50% drop in *median* throughput for the file size range (2.2 GB - 3.1 GB) relative to the throughput levels for files smaller than 2.2 GB or larger than 3.1 GB. All that can be said here is that the number of stripes and number of streams are not the causes for this drop, but other factors such as disk I/O usage and CPU usage at the servers, and network link utilization, could have caused this drop in median throughput. The second observation is that the median throughput is lower for the 8-stream group when compared to the median throughput for the 1-stream group for this file size range. An examination of sample sizes for each file size bin shows that for the (2.1-2.2 GB) bin, the number of observations for the 1-stream group is still quite large at 618 (see Fig. 5). But for sizes larger than 2.3 GB, the number of transfers in the 1-stream group for each file size bin may be too small (less than 300), and hence the median throughput values may not be representative.

The third and most relevant point to note in Fig. 4 is that the throughput is roughly the same for both 1-stream and 8-stream groups for large file sizes. This indicates that packet losses are rare if any, because if TCP packet losses occur, TCP congestion control algorithms will cause cwnd to drop. Such a drop will cause the throughput of 1-stream transfers

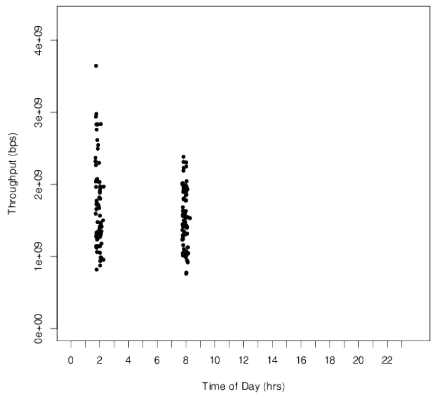

Fig. 6: Throughput of the 145 32GB NERSC-ORNL transfers as a function of time of day

to be lower than that of 8-stream transfers. Since this is not observed, our hypothesis is that there are few packet losses if any. We plan to test this hypothesis using tstat [23], a tool that reports packet loss information on a per-TCP connection basis.

In summary, the number of streams is an important factor in high BDP paths for small files, but not for large files.

C. Impact of time-of-day and link utilization

In each of the 145 32 GB NERSC-ORNL transfers, the number of stripes used was 1, and the number of parallel TCP streams used was 8. Therefore, these factors do not contribute to the variance.

Time-of-day dependence Since the start time of each transfer is logged, this was used to examine whether there was a time-of-day dependence of the throughput for the 32 GB transfers. The results are shown in Fig. 6. All the 32 GB transfers started at either 2 AM or 8 AM. Some of the transfers at 2 AM appear to have received higher levels of throughput, but there is significant variance within each set.

Dependence on link utilization First, the `traceroute` tool was used to determine the path taken by packets in these 32 GB transfers. Next, Simple Network Management Protocol (SNMP) link usage measurements were obtained for the ESnet portion of the path of the 32GB transfers[1]. ESnet configures its routers to collect byte counts (incoming and outgoing) on all interfaces on a 30 second basis. SNMP byte counts were obtained from the egress interfaces on the path of the 32GB transfers. Since the GridFTP logs include both STORE operations (files moved from ORNL to NERSC) and RETR operations (files moved from NERSC to ORNL), the appropriate interfaces were used for each GridFTP transfer.

The start and end times of the GridFTP transfers will typically not align with the 30-sec SNMP time bins. For example, Table X shows the SNMP byte counts reported for one of the interfaces on the path for each of the 30-sec bins within the duration of one of the 32GB transfers.

To determine the average traffic load on each link L of the path during a 32GB transfer, the following method was used. Let $\tau_{i1}, \tau_{i2}, \cdots \tau_{im}$ represent the 30-sec SNMP time bin boundaries such that $\tau_{i1} \leq s_i$, where s_i represents the start time of the i^{th} GridFTP transfer, and $\tau_{im} \geq (s_i + D_i)$, where D_i is the duration of the i^{th} GridFTP transfer. Indexing on link L is omitted for clarity. Assume that the SNMP byte counts for link L for these time bins with start times $\tau_{i1}, \tau_{i2}, \cdots, \tau_{i(m-1)}$ are correspondingly $b_1, b_2, \cdots, b_{m-1}$. The total number of bytes transferred on link L during the i^{th} GridFTP transfer is computed as follows:

$$B_i = b_1 \frac{(\tau_{i2} - s_i)}{30} + (\sum_{j=2}^{m-2} b_j) + b_{m-1} \frac{(s_i + D_i - \tau_{i(m-1)})}{30} \quad (1)$$

The 32 GB GridFTP transfers were divided into four quartiles based on throughput. Correlation coefficients between the GridFTP bytes and the total SNMP reported bytes B_i estimated to have been transferred during the i^{th} GridFTP transfer, from five routers (`rt1` through `rt5`) were obtained for each of these quartiles, and for the whole set of 145 transfers. Results are shown in Table XI. The high correlations between the GridFTP byte counts and SNMP byte counts suggest that the 32GB transfers dominated the total traffic on the ESnet links. While this was to be expected for the highest quartile, the results for the smallest quartile, which shows the 32 GB transfers dominating the total number of bytes, is surprising. One possible explanation is that this is just a single sample, and in general, correlation coefficients are likely to be smaller for the lower quartiles.

Also correlations between GridFTP transfer sizes ($t_i \times D_i$) (where t_i is the i^{th} transfer throughput) and the remaining traffic ($B_i - (t_i \times D_i)$) were also computed, and are shown for the multiple links on the path in Table XII. The low correlations imply that the remaining traffic does not effect GridFTP transfer throughput. Characteristics of link bandwidth usage averaged over the duration of each GridFTP transfer (B_i/D_i) across the 145 transfers is shown in Table XIII, which shows that even the maximum loads are only slightly more than half the link capacities (which are all 10 Gbps).

[1]SNMP data for 2 out of the 7 routers on the ESNet portion of the path were unavailable for the duration of interest.

TABLE XI: Correlation between GridFTP bytes and total number of bytes B_i (NERSC-ORNL)

	rt1	rt2	rt3	rt4	rt5
1st Qu.	0.677	0.604	0.719	0.750	0.749
2nd Qu.	0.419	0.147	0.138	0.327	0.294
3rd Qu.	0.538	0.592	0.543	0.415	0.371
4th Qu.	0.782	0.872	0.797	0.789	0.790
All	0.902	0.922	0.919	0.918	0.918

TABLE XII: Correlation between GridFTP bytes and bytes from other flows (NERSC-ORNL)

	rt1	rt2	rt3	rt4	rt5
1st Qu.	0.254	0.188	0.429	0.505	0.486
2nd Qu.	0.269	-0.067	-0.110	0.089	0.071
3rd Qu.	0.059	0.157	0.110	0.015	-0.039
4th Qu.	0.196	0.328	0.239	0.287	0.276
All	0.351	0.365	0.443	0.524	0.527

In summary, the time-of-day factor appears to have a minor impact on transfer throughput. The link utilization analysis shows that the backbone network links are relatively lightly loaded making the impact of other traffic on the GridFTP transfers fairly small. On the other hand, the analysis shows that the science flows dominate the total traffic, which means they could have an adverse effect on delay/jitter-sensitive real-time audio/video flows. Loads on links within the NERSC and ORNL campuses will be obtained and analyzed in future work. About the access links (from NERSC and ORNL to ESnet), these are part of ESnet, because ESnet locates its own (provider-edge) routers within the NERSC and ORNL campuses. As these links are part of ESnet, SNMP link loads were available to us and have been included in the above-described analysis.

D. Impact of concurrent GridFTP transfers

This analysis considers the effect of concurrent GridFTP transfers at the NERSC GridFTP server on the throughput of a set of NERSC-ANL memory-to-memory transfers. For each of the 84 memory-to-memory transfers, the duration is divided into intervals based on the number of concurrent transfers being executed by the NERSC GridFTP server. For example, for a particular transfer, Fig. 7 shows that there were 7 concurrent transfers during the first 6.56 seconds, 6 concurrent transfers during the next 3.98 seconds, etc.

For the i^{th} transfer, a predicted throughput value $\tilde{t}_i$ is

TABLE XIII: Average link load (Gbps) during the 145 32GB transfers

	rt1	rt2	rt3	rt4	rt5
Min	0.914	0.971	0.925	0.876	0.868
1st Qu.	1.679	1.768	1.780	1.620	1.596
Median	2.073	2.145	2.114	2.000	2.065
Mean	2.187	2.250	2.268	2.187	2.185
3rd Qu.	2.592	2.665	2.633	2.560	2.629
Max	4.573	4.701	5.065	5.184	5.152

TABLE X: SNMP byte counts within the duration of an example 32GB transfer that lasted from 1283305629 (UTC) to 1283305789

	1283305620	1283305650	1283305680	1283305710	1283305740	1283305770	(Total)
SNMP byte counts	4595537584	7355952863	7178287530	7090113782	6686499265	5889479201	38795870225

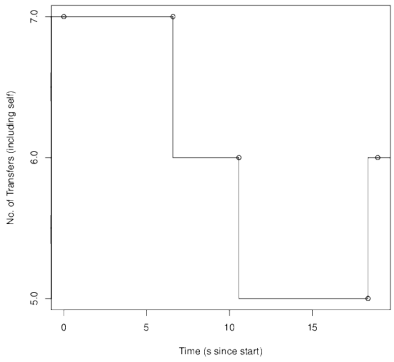

Fig. 7: No. of concurrent transfers within the duration of a particular transfer

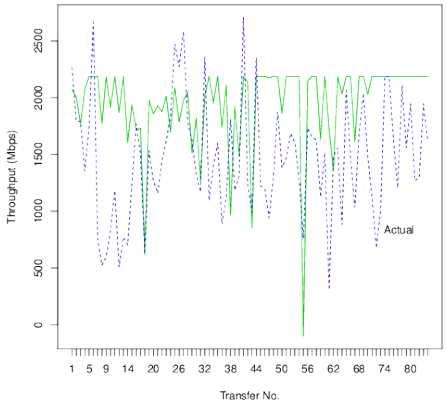

Fig. 8: Actual and predicted throughput values for memory-to-memory transfers from ANL to NERSC ($\rho = 0.2513$)

computed as follows:

$$\tilde{t_i} = \sum_{j=1}^{j_{max}} (R - \sum_{k=1}^{n_{ij}} t_k) \times \frac{d_{ij}}{D_i} = R - \sum_{j=1}^{j_{max}} \sum_{k=1}^{n_{ij}} \frac{t_k d_{ij}}{D_i} \quad (2)$$

where R is a theoretical maximum aggregated throughput that a server can support across all concurrent transfers, n_{ij} is the number of concurrent transfers in the j^{th} interval of the i^{th} transfer, t_k is the recorded throughput of the k^{th} concurrent transfer, d_{ij} is the duration of j^{th} interval of the i^{th} transfer, and D_i is the duration of the i^{th} transfer. In this formulation, R is assumed to be a constant, but the rate that a server can sustain for the i^{th} transfer depends on many time-varying factors. This makes it difficult to accurately estimate R.

Fig. 8 plots the predicted throughput values ($\tilde{t_i}$) against actual throughput values (t_i) for $1 \leq i \leq 84$, with $R = 2.19$ Gbps. The correlation coefficient, ρ, between the predicted and actual transfer throughput values, is 0.289. The choice of R impacts the predicted throughput plot, but it does not impact correlation. Here, the 90^{th} percentile of the transfer throughput from the dataset under analysis is chosen for R. A correlation analysis is also carried out on a per-quartile basis by dividing the transfers into quartiles based on actual throughput values (t_i). The correlation coefficients are 0.141, 0.051, 0.191, and 0.347, for the four quartiles, respectively.

In summary, this analysis shows us that concurrent transfers have a weak impact on transfer throughput.

VIII. Conclusions

This work presented an analysis of GridFTP transfer logs for four paths, with a focus of determining the usability of dynamic virtual circuits (VCs) for large dataset movement. From the data analysis, our conclusions are as follows. First, users typically transfer large numbers of files in sessions, and session sizes are large enough that even on high-rate paths, their durations will be long enough to justify the overhead of VC setup delay. Second, transfers occur at a significant fraction of link capacity, which means they can have adverse effects on real-time audio/video flows. The use of virtual circuits provides the opportunity, during the setup phase, for the network to control the path taken by scientific data transfers, and to isolate their packets into separate virtual queues. Third, packet losses appear to be rare in these research-and-education networks, a finding that can impact the design of new transport protocols for high bandwidth-delay product paths. Fourth, competition for server resources appears to be greater than for network resources, which means scheduling mechanisms are needed for server resources if transfer throughput variance is to be reduced.

IX. Acknowledgments

The authors thank Brent Draney and Jason Lee, NERSC, Jon Dugan and Joe Burrescia, ESnet, Joseph Bester and Stuart Martin, ANL, for helping us obtain the data analyzed in this work. The UVA portion was supported by NSF grants OCI-1038058, OCI-1127340, and CNS-1116081, and U.S. DOE grants DE-SC0002350 and DE-SC0007341. The ESnet portion was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. DOE under Contract No. DE-AC02-05CH11231. The NCAR portion was supported through NSF Cooperative Grant NSF01 that funds NCAR, and through the grant OCI-1127341. The ANL portion was supported by the U.S. DOE, under Contract No. DE-AC02-06CH11357. This research used resources of the ESnet ANI Testbed, which is supported by the Office of Science of the U.S. DOE under contract DE-AC02-05CH11231, funded through the American Recovery and Reinvestment Act of 2009.

REFERENCES

[1] J. T. Overpeck, G. A. Meehl, S. Bony, and D. R. Easterling, "Climate data challenges in the 21st century," *Science*, vol. 331, no. 6018, pp. 700–702, 2011. [Online]. Available: http://www.sciencemag.org/content/331/6018/700.abstract

[2] GridFTP v2 Protocol Description. [Online]. Available: http://www.ggf.org/documents/GFD.47.pdf

[3] Fast data transfer. [Online]. Available: http://monalisa.cern.ch/FDT/

[4] bbftp. [Online]. Available: http://doc.in2p3.fr/bbftp/

[5] S. Sarvotham, R. Riedi, and R. Baraniuk, "Connection-level analysis and modeling of nework traffic," in *ACM SIGCOMM Internet Measurement Workshop 2001*, November 2001, pp. 99–104.

[6] On-Demand Secure Circuits and Advance Reservation System (OSCARS). [Online]. Available: http://www.es.net/services/virtual-circuits-oscars/

[7] J. Postel and J. Reynolds, "File Transfer Protocol," RFC 959 (Standard), Internet Engineering Task Force, Oct. 1985, updated by RFCs 2228, 2640, 2773, 3659, 5797. [Online]. Available: http://www.ietf.org/rfc/rfc959.txt

[8] GridFTP. [Online]. Available: http://www.globus.org/toolkit/docs/latest-stable/gridftp/

[9] W. Allcock, J. Bresnahan, R. Kettimuthu, and M. Link, "The Globus striped GridFTP framework and server," in *Supercomputing, 2005. Proceedings of the ACM/IEEE SC 2005 Conference*, nov. 2005, p. 54.

[10] M. Veeraraghavan, M. Karol, and G. Clapp, "Optical dynamic circuit services," *Communications Magazine, IEEE*, vol. 48, no. 11, pp. 109 –117, november 2010.

[11] MRI-R2 Consortium: Development of Dynamic Network System (DYNES). [Online]. Available: http://www.internet2.edu/ion/dynes.html

[12] A. Lake, J. Vollbrecht, A. Brown, J. Zurawski, D. Robertson, M. Thompson, C. Guok, E. Chaniotakis, and T. Lehman, "Inter-domain Controller (IDC) Protocol Specification," May 2008.

[13] X. Zhu and M. Veeraraghavan, "Analysis and design of book-ahead bandwidth-sharing mechanisms," *Communications, IEEE Transactions on*, vol. 56, no. 12, pp. 2156 –2165, December 2008.

[14] C. Anglano and M. Canonico, "Performance analysis of high-performance file-transfer systems for grid applications," *Concurrency and Computation: Practice and Experience*, vol. 18, no. 8, pp. 807–816, 2006. [Online]. Available: http://dx.doi.org/10.1002/cpe.976

[15] Y. Zhu, D. Talia, A. Bassi, and P. Massonet, "Evaluation for high volume data transfer mechanisms in grids," Tech. Rep., October 14 2008, CoreGRID Technical Report, Number TR-0178. [Online]. Available: http://www.coregrid.net/mambo/images/stories/TechnicalReports/fellow\%20tr-0178.pdf

[16] S. Zanikolas and R. Sakellariou, "A taxonomy of grid monitoring systems," *Future Generation Computer Systems*, vol. 21, no. 1, pp. 163 – 188, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X04001190

[17] B. Balis, M. Bubak, W. Funika, T. Szepieniec, and R. Wismüller, "Monitoring and performance analysis of grid applications," in *Proceedings of the 1st international conference on Computational science: PartI*, ser. ICCS'03. Berlin, Heidelberg: Springer-Verlag, 2003, pp. 214–224. [Online]. Available: http://dl.acm.org/citation.cfm?id=1764172.1764199

[18] Kun-chan Lan and John Heidemann, "A measurement study of correlations of internet flow characteristics," *Computer Networks*, vol. 50, no. 1, pp. 46–62, 2006.

[19] Z. Yan, C. Tracy, and M. Veeraraghavan, "A hybrid network traffic engineering system," in *Proc. of IEEE 13th High Performance Switching and Routing (HPSR) 2012*, June 24-27 2012.

[20] The Lambda Station Project. [Online]. Available: http://www.lambdastation.org/

[21] B. Allen, J. Bresnahan, L. Childers, I. Foster, G. Kandaswamy, R. Kettimuthu, J. Kordas, M. Link, S. Martin, K. Pickett, and S. Tuecke, "Software as a service for data scientists," *Communications of the ACM*, vol. 55, no. 2, Feb. 2012.

[22] H. Wang, M. Veeraraghavan, R. Karri, and T. Li, "Design of a high-performance RSVP-TE hardware signaling accelerator," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 8, pp. 1588 – 1595, aug. 2005.

[23] M. Mellia, A. Carpani, and R. Lo Cigno, "Tstat: TCP statistic and analysis tool," in *Quality of Service in Multiservice IP Networks*, ser. Lecture Notes in Computer Science, M. Marsan, G. Corazza, M. Listanti, and A. Roveri, Eds. Springer Berlin / Heidelberg, 2003, vol. 2601, pp. 145–157, 10.1007/3-540-36480-3_11. [Online]. Available: http://dx.doi.org/10.1007/3-540-36480-3_11